Title: Scale-Dependent Coding of the Hippocampus in Relational Memory

Authors: Wei Ding[1], Yijing Lin[2], Bo Zhang[1*], Jia Liu[1*]

Affiliation:

[1]Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China

[2]Faculty of Psychology, Beijing Normal University, Beijing, China

***Correspondence**:

bo_zhang@tsinghua.edu.cn (Bo Zhang)

liujiaTHU@tsinghua.edu.cn (Jia Liu)

Abstract

Memory, woven into the very fabric of consciousness, serves as a time-traveling vessel in the mind, using detailed recollections not just for nostalgic reflection but as an abstract map for charting unknown futures. Here, we employed fMRI to investigate how the hippocampus (HPC) encodes detailed experiences into abstract knowledge (i.e., formation) and uses this knowledge for decision making (i.e., utilization) when human participants learned and then utilized spatial-temporal relations in a relational memory task. We found a functional gradient along the anterior-posterior axis of the HPC, characterized by representational similarity and functional connectivity with the autobiographical network. Here, the posterior HPC was more actively engaged in memory formation, whereas the anterior HPC was predominantly involved in memory utilization. Our computational modeling of relational memory further established a causal link between this functional gradient and the HPC's well-documented anatomical gradient, as optimal task performance arose from a combination of a fine-grained representation of past experiences by the posterior HPC and a coarser representation of abstract knowledge for future planning by the anterior HPC. This scale-dependent coding scheme led to the emergence of grid-like, heading direction-like, and place-like units in our neural network model, analogous to those discovered in biological brains. Taken together, our study revealed the HPC's functional gradient in representing relational memory, and further connected it to the anatomical gradient of place cells, supporting a unified framework where both spatial and episodic memory rely on relational representations that integrate spatial localization with temporal continuity.

Keywords: Hippocampus; Relational Memory; Multi-Scale Representation, fMRI, Deep Learning

Introduction

The core function of episodic memory is to remember the past and predict the future (Schacter & Addis, 2007; Addis et al., 2007; Addis et al., 2009; Schacter et al., 2012; Schacter & Madore, 2016; Gershman, 2017; Josselyn & Tonegawa, 2020; Namboodiri & Stuber, 2021). This process apparently relies on different levels of precision during encoding and retrieval stages. Encoding is often tasked with meticulously recording intricate details, necessitating fine-grained representational capacity to capture the nuanced differentiation of individual experiences (Wixted et al., 2018; Borders et al., 2022). In contrast, the retrieval process, especially for decision-making, tends to utilize a more generalized and abstracted memory schema, as this coarse form of recall provides the essential framework required to navigate through decision-making processes without the encumbrance of excessive detail (Evensmoen et al., 2015; Audrain & McAndrews, 2022). However, it is unclear how the brain accomplishes this fascinating interplay between the granular encoding and the abstracted retrieval that underscores the versatility and sophistication of our memory systems.

Recently, a unified theoretical framework has been proposed to associate episodic memory with spatial memory in the hippocampus (HPC) (Cohen & Eichenbaum, 1993; Eichenbaum, 2000; Eichenbaum & Cohen, 2014), evidenced by the abstract cognitive map (Park et al., 2020), cognitive graph (Peer et al., 2021), schema (Farzanfar et al., 2023), and geometry (Nieh et al., 2021). This framework offers a novel perspective on the HPCs role in memory encoding and retrieval by integrating episodic and spatial memory into a cohesive concept of relational memory. Interestingly, a representational gradient of the HPC in spatial scale along its anterior-posterior axis has been identified (Kjelstrup et al., 2008; Brun et al., 2008; Peer et al., 2019). Specifically, place cells in the anterior HPC of primates (or ventral HPC of rodents) show larger receptive fields than those in the posterior HPC (or dorsal HPC of rodents) (Poppenk et al., 2013; Collin et al., 2015; Robinson et al., 2015; Brunec et al., 2018). In addition, functional divisions of labor along this

representational gradient have also been identified. The anterior HPC is more engaged during autobiographical memory retrieval (Cabeza & St Jacques, 2007; McCormick et al., 2015; Mazzoni et al., 2019). In contrast, the posterior HPC is preferentially implicated in both short-term (Fernandez et al., 1998; Giovanello et al., 2009; Woolley et al., 2015; Zeithamova & Bowman, 2020) and long-term learning (Maguire et al., 2000; Weisberg et al., 2019; Weisberg & Ekstrom, 2021). Considering these anatomical and functional divisions, and the different precision levels required between encoding and retrieval stages, we hypothesize that the posterior HPC, with its smaller receptive fields and therefore higher precision, may be the site for fine-grained encoding of detailed experiences into abstract representations (i.e., abstract knowledge). In contrast, retrieval of this abstract knowledge may occur in the anterior HPC, where the larger receptive fields imply a lower precision.

To test this hypothesis, we utilized a relational memory (RM) task (Insa-Cabrera et al., 2011; Tartaglia et al., 2017), where human participants were tasked with mentally constructing and utilizing a sequence of spatial positions to predict subsequent positions (Fig. 1). In the task, participants learned the temporal structure of spatial positions through feedback on random guesses in the beginning steps of each trial (i.e., the relation formation state). Subsequently, they applied this learned sequence to accurately predict each position in the sequence in the later steps of each trial (i.e., the utilization state). Therefore, this task design allowed separate analysis of neural representations during memory formation and utilization states. In line with our hypothesis, we found greater involvement of the posterior HPC in the formation state, while the anterior HPC was more engaged in the utilization state, aligning with the representational gradient along the anterior-posterior axis of the HPC. To delve deeper into the underlying mechanism, we developed a hybrid neural network model combining convolutional and Long Short-Term Memory (LSTM) components to examine the impact of spatial scale on the formation and utilization states of memory. Computational modeling established a causal link that the relational memory's effectiveness hinged on fine-grained encoding and coarse retrieval representations. In short, our study suggests that the HPC exhibits multi-scale

representations along the anterior-posterior axis tailored to computational demands of memory at different stages.

Results

Memory formation and utilization states in relational memory task

A series of numeric strings, each composed of digits drawn from the set [1, 2, 3] and varying in length from 2 to 8, was constructed to represent sequences of locations for *Good* cues (denoted by a star) and *Bad* cues (denoted by a cross) within a single iteration, which were allocated to three horizontally aligned boxes on the screen (Fig. 1A). In a trial, sequences for both *Good* and *Bad* cues, derived from different numeric strings but of equal length, were iteratively presented until reaching a total length of 30 steps, respectively (Fig. 1B). Participants were instructed to place a *Self* marker (denoted by an avatar) at the anticipated location of the next *Good* cue while avoiding the location of the *Bad* cue, utilizing feedback from preceding steps. Participants' performance was assessed by calculating the number of correctly placing the *Self* marker in the box containing the *Good* cue (i.e., reward) minus the number of incorrectly placing in the box with the *Bad* cue (i.e., punishment). Through this process, participants first learned the spatiotemporal relations of both *Good* and *Bad* cues, then applied this knowledge to predict the locations of the next *Good* and *Bad* cues. Participants' behavioral performance exhibited a characteristic piecewise-like progression, initially manifesting a rapid rise in accuracy from random guessing, followed by a plateau of correct responses (Fig. 1C). This behavioral pattern was observed in all tested lengths of numeric strings, though a noticeable decrease in accuracy and a delay in the transition point were observed with increasing length of the numeric strings (Fig. 1D). To quantify the observation, we evaluated the fitness of a piecewise linear function to participants' behavioral performance, against various functions such as logistic, quadratic, logarithmic, and linear functions, using the coefficient of determination (R^2) as the metric. Consistent with the observation, the piecewise linear function showed the highest fitness to the behavioral performance

(Fig. 1E; M = 0.96, SD = 0.26; $p < 0.05$, corrected for multiple comparisons using permutation) compared to the alternative functions (logistic: [M = 0.90, SD = 0.04]; quadratic: [M = 0.90, SD = 0.03]; logarithmic function [M = 0.87, SD = 0.04]; and linear function: [M = 0.64, SD = 0.06]). This piecewise pattern suggests the existence of two distinct memory states with the task: one primarily pertaining to memory formation and the other to its subsequent utilization.

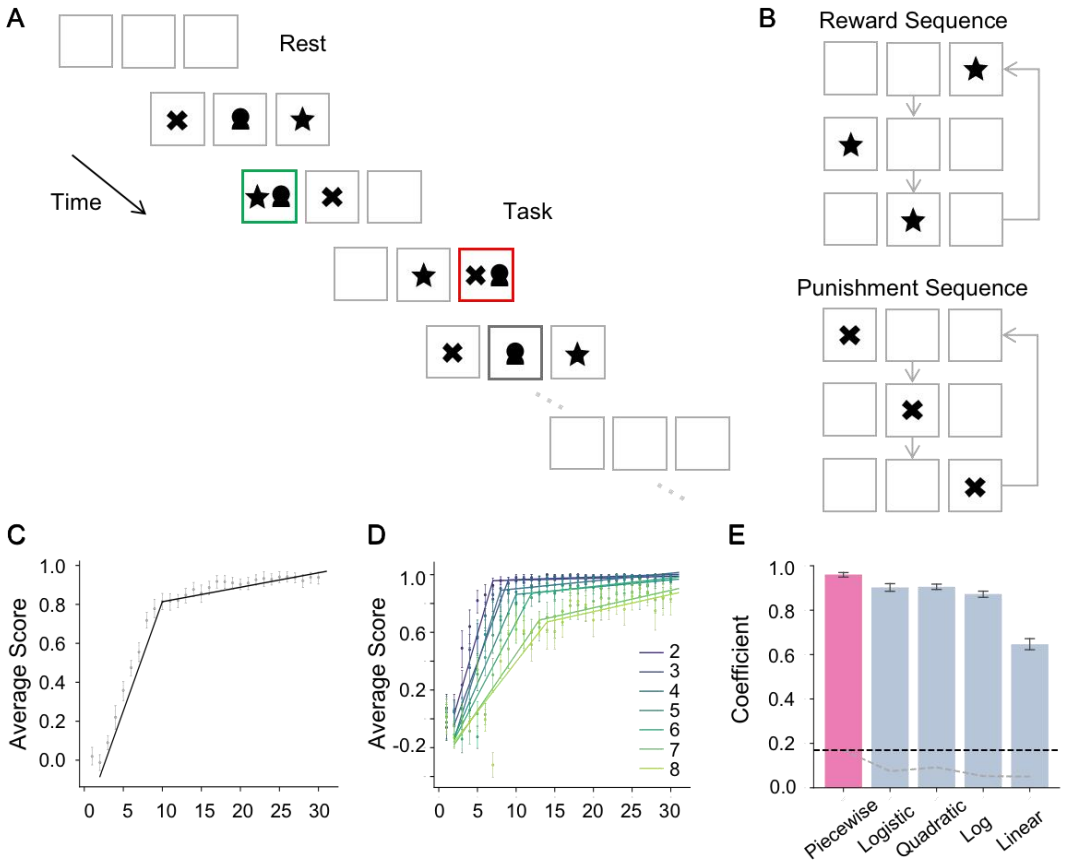

Figure 1. Experimental paradigm and behavioral performance. (A) Experimental paradigm of the RM task. A reward was presented when correctly placing the *Self* marker (avatar) in the same box (green) with the next *Good* cue (star), while a punishment was given when incorrectly placing the *Self* marker in the same box (red) with the next *Bad* cue (cross). Participants were encouraged to maximize reward and avoid punishment. (B) Exemplar sequences for the *Good* and *Bad* cues. The length of numeric strings is 3. The *Good* cue shows a [Right, Left, Middle] spatiotemporal sequence, while the *Bad* cue shows a [Left, Middle, Right] sequence. (C) Behavioral performance. Behavioral performance, averaged across sequences and participants, was improved as a function of increased steps. (D) The piecewise-like pattern, indexed by solid lines, was observed in all tested lengths of numeric strings. (E) The highest fitness to the behavioral performance was the piecewise linear function, as compared to the logistic, quadratic, logarithmic and linear functions. The dashed grey line denotes the uncorrected threshold of permutation, while the dashed black

line denotes the corrected threshold with multiple comparisons. Error bars represent standard errors across participants.

Neural correlates of memory formation and utilization

Upon identifying two distinct memory states in the RM task, we further replicated previous research (Sullivan Giovanello et al., 2004; Olson et al., 2006; Hannula et al., 2006; Giovanello et al., 2009; Schwarb et al., 2016; Zheng et al., 2024) demonstrating hippocampal activation in relational integration using representational similarity analysis (RSA; Kriegeskorte et al., 2006) and functional connectivity (FC) approaches. In the RSA analysis, we calculated the multivoxel patterns of HPC activity during the RM task for each trial and constructed a correlation coefficient matrix across trials for each voxel (see Methods). We found that all subregions of the HPC, including Dentate Gyrus, Cornu Ammonis, and Subiculum, showed significantly higher representational similarity in comparisons to the resting state (Fig. S1; initial threshold: $p < 0.01$, corrected for multiple comparisons by permutation). In the FC analysis, we designated the HPC as a seed region and computed Pearson's correlation coefficients between its averaged time course and that of other brain regions during the RM task (see Methods). We found significantly enhanced FC between the HPC and the autobiographical network (Svoboda et al., 2006; Burianova et al., 2010), including the Medial Prefrontal Cortex, Posterior Cingulate Cortex, and Angular Gyrus, during the task state in comparison to the resting state (Fig. S2; initial threshold: $p < 0.01$, corrected for multiple comparisons by permutation). Taken together, these findings align with previous research (Giovanello et al., 2009; Huijbers et al., 2011; Raichle, 2015; Smallwood et al., 2021), confirming the role of the HPC in relational memory processing during the RM task.

To investigate the neural representations of the HPC associated with the formation and utilization of relational memory, the BOLD signal series of the RM task was segregated into the formation and utilization states, which were approximately demarcated to have equal temporal durations (see Methods). An RSA analysis was performed to extract the neural representation profiles for the resting, formation, and utilization states, respectively (Fig. 2A). As expected, the

representational similarity in the HPC was significantly higher in both the formation and utilization states than that in the resting state (Fig. 2B & E, initial threshold of p = 0.01, corrected for multiple comparisons by permutation).

Interestingly, the peak of the representational similarity in the formation state was located at the posterior HPC (MNI coordinates = [57, 45, 36], $t(15) = 6.11$, $p < 0.001$, two-tailed), while in the utilization state, it was identified in the anterior HPC (MNI coordinates = [58, 57, 26]; $t(15) = 4.05$, $p < 0.001$, two-tailed). This suggests an association between the functional dissociation in memory states and the anatomical gradient in the HPC. To quantify this observation, we averaged the representational similarity along the y axis of MNI coordinate system, which produced a vector demonstrating the variation in representational similarity along the hippocampal long axis for each memory state. First, a reverse pattern was observed between the two memory states, with higher representational similarity of memory formation in the posterior HPC and for memory utilization in the anterior HPC (Fig. 2C & F; initial threshold p = 0.01, corrected for multiple comparisons by permutation). Second, we partitioned the HPC into three equal volumetric portions along its long axis (Tompary & Davachi, 2017) and then performed an ROI analysis to explore the relationship between memory state and hippocampal gradient. We found that the posterior HPC showed significantly higher representational similarity as compared to the anterior HPC during memory formation (Fig. 2D; $t(15) = 2.59$, $p < 0.05$, two-tailed). Conversely, the anterior HPC showed significantly higher representational similarity as compared to the posterior HPC during memory utilization (Fig. 2G; $t(15) = 2.66$, $p < 0.05$, two-tailed). These findings illustrated a functional division of labor along hippocampal gradient for the two memory states in relational memory, suggesting that memory formation particularly relies on the fine-grained representation in the posterior HPC, while memory utilization preferentially involves the coarser representation in the anterior HPC.

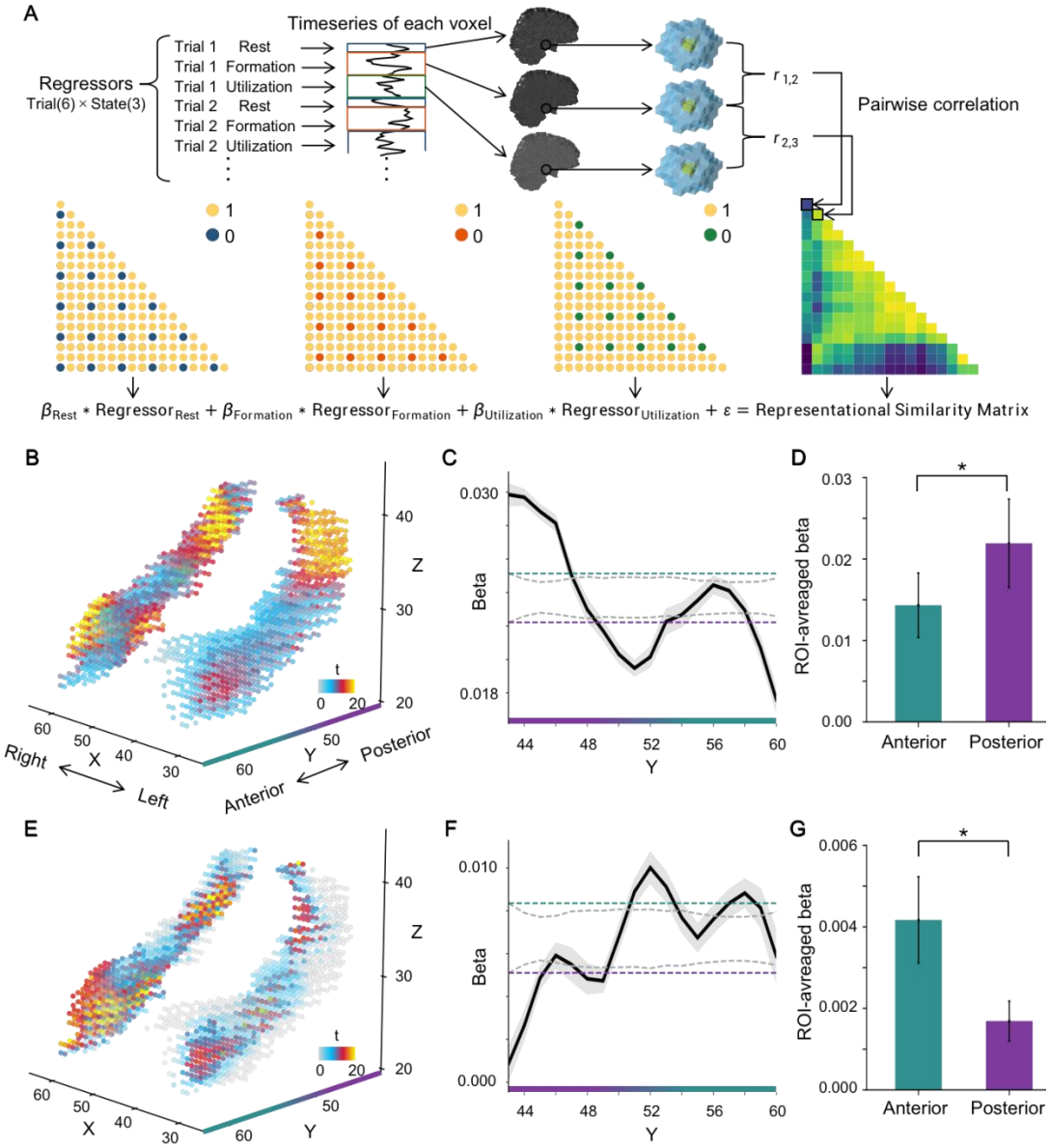

Figure 2. Function gradient in the HPC. (A) Schematic overview of RSA analysis. Multi-voxel patterns of the resting, formation and utilization states were extracted by a general linear model (GLM). The similarity matrix was constructed using a searchlight-based Pearson's correlation of the multi-voxel patterns among trials. This matrix was fitted with another GLM with states specified by binary regressors. Within the binary matrix, blue, brown, and green points denote the "same" state. The estimated parameters of the GLM were then sent to the second-level analyses. (B) Neural representation of formation state in the HPC. Each point represents a voxel in the left and right hippocampus, respectively. The value at each point, indexed by t-value, reflects the contrast in representational similarity from the memory formation relative to the resting state. The

X-, Y-, and Z-axes denote the MNI coordinates of the HPC. Specifically, the Y-axis corresponds to the hippocampal long axis, with green and purple colors representing the anterior and posterior HPC, respectively. (C) Variation in representational similarity along the hippocampal long axis during memory formation. The black line represents the averaged similarity across participants, with the shaded area denoting standard error. Gray dotted lines indicate the uncorrected thresholds derived from the 5th and 95th percentiles of the permutation distribution at each coordinate of the Y-axis, whereas green and purple dotted lines indicate the corrected threshold defined by the maximum of uncorrected thresholds. (D) ROI analysis of representational similarity between the anterior and posterior HPC. The representational similarity of memory formation state was significantly higher in the posterior HPC than that in the anterior HPC. (E – G) Neural representation of memory utilization state in the HPC. *: $p < 0.05$; Error bars: standard error.

Given the aforementioned evidence of the HPC's division in representational similarity and its FC with the autobiographical network during the MR task, we predicted that the posterior HPC should predominantly connect with the autobiographical network during the formation state, while the anterior HPC should show stronger FC with this network during the utilization state.

To test this conjecture, we analyzed the FC between the HPC and the autobiographical network during the formation and utilization states of the RM task, respectively. Note that the medial temporal lobe was excluded from this analysis to avoid overlap with the HPC. We found that the posterior HPC showed stronger FC with the autobiographical network during the formation state (Fig. 3A; initial threshold $p = 0.01$, corrected for multiple comparisons by permutation), with the FC peak located at the coordinate [55, 52, 27] ($t(15) = 4.98$, $p < 0.001$, two-tailed). Conversely, during the utilization state, stronger FC with the autobiographical network was found in the anterior HPC (Fig. 3B; initial threshold $p = 0.01$, corrected for multiple comparisons by permutation), with the peak at [57, 61, 23] ($t(15) = 3.29$, $p < 0.001$, two-tailed). To quantify this observation, a contrast analysis between the utilization and formation states was performed on FC along the hippocampal long axis. We found that the FC difference between the two states shifted from negative to positive along the axis, moving from the posterior to anterior HPC (Fig. 3C; initial

threshold p = 0.01, corrected for multiple comparisons by permutation). This connectivity gradient along the hippocampal long axis was further confirmed by a significant two-way interaction between the memory state (formation versus utilization) and the hippocampal gradient (posterior versus anterior) by an ROI analysis (Fig. 3D; $F(1, 15) = 9.03$, $p = 0.004$). *Post-hoc* t-tests showed that in the anterior HPC, the FC during the utilization state was significantly higher than that of the formation state ($t(15) = 5.44$, $p < 0.001$, two-tailed), while a reverse pattern, though not significant, was found in the posterior HPC ($t(15) = 1.55$, $p = 0.14$, two-tailed). In summary, the FC analysis also reveals a connectivity gradient of the HPC with the autobiographical network during the RM task.

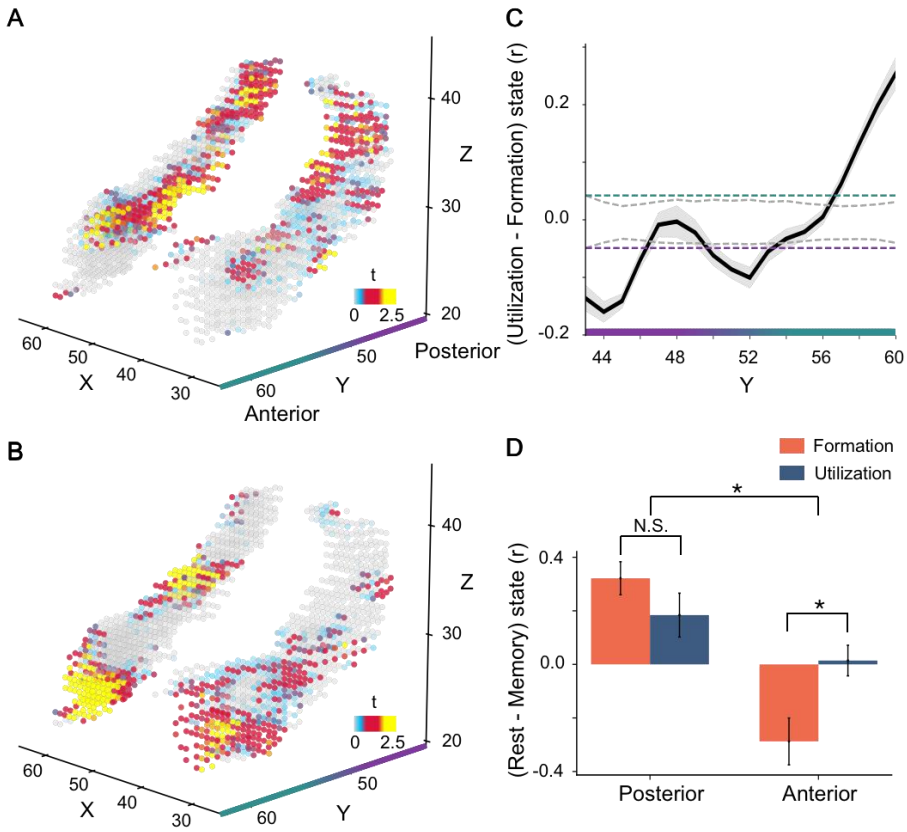

Figure 3. Connectivity gradient in the HPC. (A) Difference in FC of the formation state relative to the utilization state. Pearson's correlation was performed between the mean BOLD signal series within the autobiographic network and that in each voxel of the HPC for the formation and utilization states, respectively. The difference in FC was indexed by t-value. (B) Difference in FC of the utilization state relative to formation state. (C) Difference in FC of the utilization state

relative to the formation state along the hippocampal long axis. (D) FC strength of memory state (formation versus utilization) and hippocampal gradient (posterior versus anterior) compared to the rest state revealed a significant interaction. *: $p < 0.05$; Error bars: standard error.

The causal link between functional and anatomical gradients

The findings from both representational similarity and functional connectivity analyses suggest that the posterior HPC is predominantly involved in memory formation and the anterior HPC in memory utilization. Interestingly, these function and connectivity gradients coincide with the anatomical gradient of the HPC, where the posterior HPC is characterized by a higher density of place cells with smaller receptive fields, and the anterior HPC by place cells with larger receptive fields. Specifically, previous studies have shown that the posterior HPC is preferentially associated to both short- and long-term learning (Fernandez et al., 1998; Giovanello et al., 2009; Woolley et al., 2015; Zeithamova & Bowman, 2020; Maguire et al., 2000; Weisberg et al., 2019; Weisberg & Ekstrom, 2021), implying its role in encoding information into an abstract, fine-grained representation. In contrast, the anterior HPC is often co-activated with visual sensory regions during autobiographical memory retrieval (Cabeza & St Jacques, 2007; McCormick et al., 2015; Mazzoni et al., 2019), implying its involvement in computational processes of visual action at a coarse precision. Therefore, if there is a causal anatomical-functional link, we would expect the memory formation to require a fine-grained precision, and the memory utilization to favor a coarser precision.

To test this hypothesis, we developed a hybrid Convolutional Long Short-Term Memory (C-LSTM) neural network to simulate how information is encoded and then retrieved during the RM task (Fig. 4A). The C-LSTM comprises three components: (1) a convolutional layer for encoding the visual presentation of *Good* and *Bad* cue sequences, (2) an LSTM layer for abstractly storing the spatiotemporal locations of these cues, and (3) a transposed convolution layer for predicting the location of the next *Good* cue through *Self* marker placement. A series of the C-LSTMs, varying the kernel sizes of the convolutional layers (representing visual-action resolution, VR)

and the hidden sizes of the LSTM layer (representing memory resolution, MR), were trained. In this context, the MR parameter corresponds to the resolution necessary for memory formation, and the VR parameter to the resolution required by memory utilization. Our aim was to determine which combinations of VR and MR yield optimal performance of the C-LSTM in the RM task.

Computational simulations revealed that both the MR and VR parameters significantly modulated C-LSTM's performance, which was characterized by the prediction accuracy of *Good* cue location. Specifically, the performance increased with the increase of the MR parameter ($F(14, 195) = 13.56$, $p < 0.001$) and as the decrease of the VR parameter ($F(14, 195) = 4.39$, $p < 0.001$). Additionally, the interaction between MR and VR was also significant ($F(1, 195) = 1.31$, $p < 0.05$), suggesting that the increase of MR elevated the slope of the influence of VR on performance. Therefore, optimal performance was reached by a C-LSTM with a high MR value combined with a low VR value (Fig. 4B), aligning with our prediction that memory formation requires a fine-grained representation, while memory utilization prefers a coarse representation.

To further unveil how spatiotemporal information was encoded and then retrieved in our model, we extracted the feature maps of each single unit in the transposed convolution layer, because their responses directly determine model's decision. For each unit, its weight was retained while the weights of all other units were set to zero. Accordingly, the output of the deconvolutional layer yielded a feature map of this single unit. Interestingly, the feature maps of deconvolutional units embedded the spatial structure of visual stimuli in either the vertical dimension (Fig. 4C, top), the horizontal dimension (Fig. 4C, bottom), or both the vertical and horizontal dimensions (Fig. 4C, middle). These hierarchical activity patterns of neural units are analogous to the place cells, head direction (HD) cells, and grid cells as observed in biological brains (O'Keefe & Conway, 1978; Taube et al., 1990; Hafting et al., 2005). Specifically, place-like units showed selective activation for the horizontal target locations, HD-like units responded with a band-like pattern across the rows vertically across the space, and grid-like units exhibited the square-like activation patterns

tessellating the entire space (Fig. 4C). To quantitatively identify these three types of units in the transposed convolution layer and their roles on model decision, we first applied Fast Fourier Transform (FFT) to identify the spatial frequency of the feature map of each unit from 225 C-LSTMs (i.e., all possible pairs of 15 VRs with 15 MRs). Specifically, grid-like units are characterized by the minimum spectral power at 3-cycle from horizontal and vertical axes, HD-like units was expected to show the spectral power at 1-cycle in the vertical axis, and place-like units would exhibit the spectral power at 1-cycle in horizontal axes. Second, we equally partitioned C-LSTMs into three group (i.e., ①②③) according to their performance varied along the diagonal of VR-MR matrix (Fig. 4B). In this way, the participation of the three types of units can be accessed by the group of models from poor to optimal performance. As the result, significant positive correlations were revealed for the number of either grid-like or place-like units with model performance (Fig. 4D; grid-like unit: $r = 0.38$, $p < 0.001$; place-like unit: $r = 0.71$, $p < 0.001$). Correspondingly, the FFT power of these two types of unit significantly increased with model performance (Fig. 4E; grid-like unit: $F(2, 132) = 70.02$, $p < 0.001$; place-like unit: $F(2, 132) = 177.72$, $p < 0.001$). However, no significant variation of the FFT power was found for HD-like units among the models with diverse range of performance (Fig. 4D & E; $r = -0.019$, $p > 0.05$; ANOVA: $F(2, 132) = 1.26$, $p = 0.29$). Therefore, the HD-like units appeared consistently across models, regardless of performance (Fig. 4E, left); however, the presence of grid- and place-like units decreased in models with medium performance and was absent in those with poor performance (Fig. 4E, middle & right). This pattern suggests a hierarchical emergence in relational memory processing, with HD-like units appearing first, followed by grid- and place-like units (see also [Wills & Cacucci, 2014] for similar patterns in mice during the development).

Taken together, our computational model demonstrates a causal link between the multi-scale representations in the HPC and the cognitive states of relational memory. Importantly, the emergence of grid-like, HD-like, and place-like units in the C-LSTM implies the essential role of the HPC in processing relational memory, aligning with

observed biological phenomena (O'Keefe & Conway, 1978; Taube et al., 1990; Hafting et al., 2005).

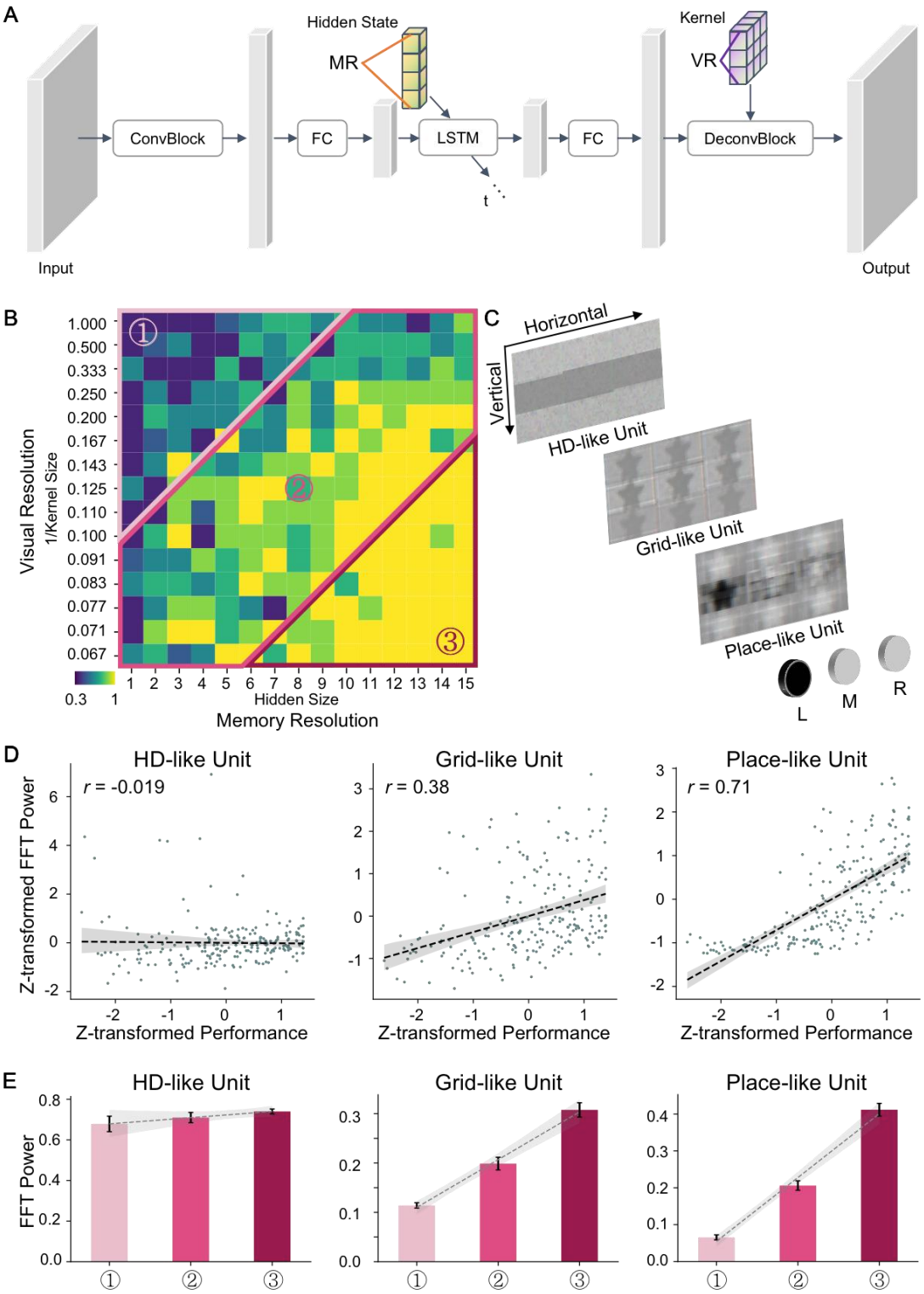

Figure 4. Computational modeling of memory formation and utilization. (A) Architecture of the hybrid neural network (C-LSTM). The C-LSTM comprises convolutional layers with an

LSTM. The input to the C-LSTM is the visual stimuli from the RM task, and the output is an image predicting the location of the next *Good* cue (denoted by a star). ConvBlock, FC, and DconvBlock denotes convolution layer, fully connected layer, and deconvolutional layer, respectively. MR denotes the size of the hidden state in the LSTM, while VR denotes the kernel size of the deconvolutional layers. (B) Behavioral Performance of C-LSTM. The performance varied significantly across combinations of MR and VR. Each cell in the matrix indicates model performance for a specific pair of VR and MR, color-coded for prediction accuracy (dark blue: poor; light yellow: optimal). Models outlined in color were chosen for further analyses in (E). (C) Exemplar feature maps from deconvolutional layer units. The feature maps of these units were analogous to the activation patterns of biological head direction cells (top), grid cells (middle), and place cells (bottom). (D) Pearson correlations between FFT power and model performance across unit types. (E) Averaged FFT power across unit types. The presence of the HD-like, grid-like, and place-like units in models with poor ①, medium ②, and optimal ③ performance. Dashed black line denotes regression line. Shaded gray area denotes SE. Error bars: standard error.

Discussion

The hippocampus acts as a bridge between our past and future, enabling us to learn from previous experiences and utilize that knowledge to navigate future challenges (Schacter & Addis, 2007; Addis et al., 2007; Brown et al., 2016; Gershman, 2017; Josselyn & Tonegawa, 2020; Namboodiri & Stuber, 2021). In this study, we used fMRI to investigate how the hippocampal multi-scale representation bridges the two key states of relational memory: formation and utilization. Our study revealed a functional gradient along the anterior-posterior axis of the HPC, characterized by representational similarity and functional connectivity with the autobiographical network. Specifically, the posterior HPC was predominantly involved in memory formation, while the anterior HPC was more actively engaged in memory utilization. Computational modeling of relational memory further established a causal link between the functional gradient and the HPC's well-documented anatomical gradient, as evidenced by the size of place fields progressively decreasing along the anterior-posterior axis (Kjelstrup et al., 2008; Brun et al., 2008; Brunec et al., 2018).

That is, optimal performance in relational memory tasks likely derived from the fine-grained representation of past experiences by the posterior HPC and the coarser representation of abstract knowledge for future planning by the anterior HPC. Taken together, our study not only discovered the HPC's functional gradient for representing relational memory but also demonstrated its origin from the anatomical gradient of place cells, hereby echoing a fundamental principle in neuroscience that structure determines functionality (Suárez et al., 2020; Bassett & Sporns, 2017; Babaeeghazvini et al., 2021).

The functional division of labor in the HPC has been extensively studied, evidenced by the involvement of partial hippocampal volumes in a specific memory task. For example, the retrieval of autobiographical memory selectively activated the anterior HPC (Cabeza & St Jacques, 2007; Giovanello et al., 2009; Huijbers et al., 2011; Raichle, 2015; McCormick et al., 2015; Mazzoni et al., 2019; Smallwood et al., 2021), whereas in long-term learning, only the posterior HPC showed functional and structural alterations in response to the encoding of spatial knowledge (Maguire et al., 1997; Maguire et al., 2000; Maguire et al., 2006; Weisberg et al., 2019; Weisberg & Ekstrom, 2021). Our study further extended these findings by incorporating both memory formation and utilization in one experimental paradigm, where human participants learned spatiotemporal sequences (i.e., memory formation) and then utilized the learned relations for decision making (i.e., memory utilization). This paradigm is proved efficient to study these two states of relational memory, as evidenced by the behavioral performance displaying a piecewise-like pattern with a turning point nicely separating these two states. Further, with two neural markers of representational similarity within the HPC and functional connectivity between the HPC and the autobiographical network, a functional gradient of the HPC was discovered, with the posterior HPC primarily involved in memory formation and the anterior HPC preferentially engaged in memory utilization.

Our computational modeling further linked this functional gradient to the anatomical gradient, evidenced by physiological recordings showing a decrease in the size of place fields along the anterior-posterior axis of the HPC (Poppenk et al., 2013;

Collin et al., 2015; Robinson et al., 2015; Brunec et al., 2018). By directly manipulating the size of the hidden state in the LSTM (i.e., the resolution for memory formation) and the kernel size of deconvolutional layers (i.e., the resolution for memory utilization), we explored the parameter space to identify the model that yielded the best performance. This blind exploration led us to a model exhibiting optimal performance in the relational memory task, characterized by a large scale of deconvolutional kernel size (i.e., low VR) and a small scale of LSTM hidden size (i.e., high MR). This observation provides a mechanistic insight into the alignment of the functional and anatomical gradients. That is, during memory formation, smaller sizes of place fields may facilitate the construction of high-precision abstract knowledge. In contrast, during memory utilization, larger receptive fields may facilitate the prioritization of action-related features over irrelevant ones, possibly through a mechanism of biased competition (Desimone & Duncan, 1995; Reynolds et al., 1999). Therefore, the functional and anatomical gradients in the HPC apparently reflect a computational principle tied to the multi-scale representation by place cells (Poppenk et al., 2013; Collin et al., 2015; Robinson et al., 2015; Brunec et al., 2018).

Hippocampal place cells are renowned for their pivotal role in creating a cognitive map of an environment by encoding specific locations and directions during navigation (Schiller et al., 2015; Epstein et al., 2017). Our study extends these known functionalities, proposing their significant contribution to both the formation and utilization of relational memory. The integral role of place cells aligns with the concept that the HPC provides a unified framework where spatial memory and episodic memory rely on relational representations that integrate spatial localization with temporal continuity (Cohen & Eichenbaum, 1993; Eichenbaum, 2000; Eichenbaum & Cohen, 2014). Consistent with this speculation, our analysis of the feature maps of single units in the deconvolution layers revealed the presence of place-like units, as well as HD-like and grid-like units. Critically, the absence of these units, especially the place-like and grid-like units, significantly impaired the model's performance in the relational memory task. These self-emerged units in our computational model, analogous to their biological counterparts essential in

navigation, may constitute the implementational basis for relational representations. Therefore, our findings not only shed light on a broader application of hippocampal place cells in relational memory processing but also reinforce the concept of a unified framework underpinned by these specialized cells in the hippocampus (Epstein et al., 2017; Whittington et al., 2022; Sorscher et al., 2023).

There are several limitations warranting further investigation. First, our study suggests a continuous functional gradient along the anterior-posterior axis of the HPC, contrasting with the view of distinct functionalities for the posterior and anterior regions. However, this functional gradient, although in alignment with the anatomical gradient in the size of place fields, may also be influenced by our experimental paradigm. In the RM task, memory formation and utilization were intertwined; for example, elements of memory utilization were present even during the formation state due to task requirements. Similarly, in our fMRI data analysis, the BOLD signal series were bifurcated into two equal-length segments, the first half as the formation state and the second half as the utilization state, primarily because a distinct turning point at the trial level was not prominent. This intermingling nature of the formation and utilization states may contribute to the observed gradient in both representational similarity and functional connectivity. Future research is needed to more distinctly decouple these two states to better depict the functional gradient of the HPC. Second, the memory formation and utilization states in our study are markedly different from the states of retrieving the past and imagining the future, which are associated with scene construction (Gaesser et al., 2013) and engage both the anterior and posterior HPC (Strange et al., 2014). Future research should aim to characterize relational memory from multiple dimensions, such as spatial and temporal dimensions, and then integrate these dimensions to comprehensively understand relational memory in the hippocampus, hereby offering deeper insights into the multifaceted functional roles of the HPC. Third, while our model successfully simulated the hippocampal multi-scale coding at behavioral level, it is important to recognize its limitations in terms of biological plausibility, especially concerning its anatomical implementation. Future research should aim to develop models that more accurately resemble the HPC's

469 actual anatomical structure, thereby bridging the gap between computational
470 modeling and biological mechanisms of the HPC.

Acknowledgments:

The work was supported by China Postdoctoral Science Foundation (2022M710470, B.Z.), Beijing Municipal Science & Technology Commission & Administrative Commission of Zhongguancun Science Park (Z221100002722012), and Double First-Class initiative Funds for Discipline Construction of Tsinghua University (J.L.).

Author contributions:

J.L. conceived the study. Y.L. developed the task. W.D and Y.L collected the behavioral and fMRI data. W.D. performed the data analyses under B.Z.'s supervision. B.Z. developed the C-LSTM network. B.Z. and W.D. performed the network simulation. W.D., B.Z. and J.L. wrote the paper.

Competing interests:

The authors declare no competing interests.

Data availability:

The data that support the findings of the present study are available from the corresponding author (J.L.) upon request.

Code availability:

All analyses reported in the present study were conducted with the customized codes written in Python (version 3.9) and Matlab (version 2021a). The codes will be available from the corresponding author (J.L.) upon request. The neuroimaging toolbox and machine learning framework used by the present work are publicly available as following: FSL (version 6.0, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki), BrainNet Viewer (version 1.7, https://www.nitrc.org/projects/bnv/), and Torch (version 1.9, https://pytorch.org/).

References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45*(7), 1363–1377. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Addis, D. R., Pan, L., Vu, M.-A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia, 47*(11), 2222–2238. https://doi.org/10.1016/j.neuropsychologia.2008.10.026

Audrain, S., & McAndrews, M. P. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. *Nature Communications, 13*(1), 5795. https://doi.org/10.1101/2020.10.11.335166

Bartels, A., & Zeki, S. (2005). The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage, 22*(1), 419–433. https://doi.org/10.1016/j.neuroimage.2004.01.007

Birn, R. M., Diamond, J. B., Smith, M. A., & Bandettini, P. A. (2006). Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage, 31*(4), 1536–1548. https://doi.org/10.1016/j.neuroimage.2006.02.048

Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., & Moser, M. B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus, 18*(12), 1200–1212. https://doi.org/10.1002/hipo.20504

Burianova, H., McIntosh, A. R., & Grady, C. L. (2010). A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *Neuroimage, 49*(1), 865–874. https://doi.org/10.1016/j.neuroimage.2009.08.066

Brown, T., Carr, V., LaRocque, K., Favila, S., Gordon, A., Bowles, B., Bailenson, J., & Wagner, A. (2016). Prospective representation of navigational goals in the

human hippocampus. *Science, 352*, 1323 – 1326. https://doi.org/10.1126/science.aaf0784

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience, 20*(3), 353–364. https://doi.org/10.1038/nn.4502

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z. X., & Moscovitch, M. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Current Biology, 28*(13), 2129–2135. https://doi.org/10.1016/j.cub.2018.05.016

Babaeeghazvini, P., Rueda-Delgado, L. M., Gooijers, J., Swinnen, S. P., & Daffertshofer, A. (2021). Brain Structural and Functional Connectivity: A Review of Combined Works of Diffusion Magnetic Resonance Imaging and Electro-Encephalography. *Frontiers in Human Neuroscience, 15*, 721206. https://doi.org/10.3389/fnhum.2021.721206

Borders, A. A., Ranganath, C., & Yonelinas, A. P. (2022). The hippocampus supports high-precision binding in visual working memory. *Hippocampus, 32*(3), 217–230. https://doi.org/10.1002/hipo.23401

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. MIT press. https://doi.org/10.5860/choice.31-5148

Conway, M. A., Pleydell-Pearce, C. W., & Whitecross, S. E. (2001). The neuroanatomy of autobiographical memory: A slow cortical potential study of autobiographical memory retrieval. *Journal of Memory and Language, 45*(3), 493–524. https://doi.org/10.1006/jmla.2001.2781

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in cognitive sciences, 11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Collin, S. H., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature neuroscience, 18*(11), 1562–1564. https://doi.org/10.1038/nn.4138

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience, 18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Eichenbaum, H. (2000). Hippocampus: mapping or memory?. *Current biology, 10*(21), R785–R787. https://doi.org/10.1016/S0960-9822(00)00763-6

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron, 83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Evensmoen, H. R., Ladstein, J., Hansen, T. I., Møller, J. A., Witter, M. P., Nadel, L., & Håberg, A. K. (2015). From details to large scale: The representation of environmental positions follows a granularity gradient along the human hippocampal and entorhinal anterior–posterior axis. *Hippocampus, 25*(1), 119–135. https://doi.org/10.1002/hipo.22357

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience, 20*(11), 1504–1513. https://doi.org/10.1038/nn.4656

Fernandez, G., Weyerts, H., Schrader-Bölsche, M., Tendolkar, I., Smid, H. G., Tempelmann, C., ... & Heinze, H. J. (1998). Successful verbal encoding into episodic memory engages the posterior hippocampus: a parametrically analyzed functional magnetic resonance imaging study. *Journal of Neuroscience, 18*(5), 1841–1847. https://doi.org/10.1523/JNEUROSCI.18-05-01841.1998

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673–9678. https://doi.org/10.1073/PNAS.0504136102

Farzanfar, D., Spiers, H. J., Moscovitch, M., & Rosenbaum, R. S. (2023). From cognitive maps to spatial schemas. *Nature Reviews Neuroscience, 24*(2), 63–79. https://doi.org/10.1038/s41583-022-00655-9

Giovanello, K. S., Schnyer, D., & Verfaellie, M. (2009). Distinct hippocampal regions make unique contributions to relational memory. *Hippocampus, 19*(2), 111–117. https://doi.org/10.1002/hipo.20506

Gaesser, B., Spreng, R. N., McLelland, V. C., Addis, D. R., & Schacter, D. L. (2013). Imagining the future: Evidence for a hippocampal contribution to constructive processing. *Hippocampus, 23*(12), 1150–1161. https://doi.org/10.1002/hipo.22152

Gershman, S. J. (2017). Predicting the past, remembering the future. *Current Opinion in Behavioral Sciences, 17*, 7–13. https://doi.org/10.1016/j.cobeha.2017.05.025

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience, 26*(32), 8352–8359. https://doi.org/10.1523/JNEUROSCI.5222-05.2006

Huijbers, W., Pennartz, C. M., Cabeza, R., & Daselaar, S. M. (2011). The hippocampus is coupled with the default network during memory retrieval but not during memory encoding. *PLoS ONE, 6*(4), e17463. https://doi.org/10.1371/journal.pone.0017463

Insa-Cabrera, J., Dowe, D. L., Espana-Cubillo, S., Hernández-Lloreda, M. V., & Hernández-Orallo, J. (2011). Comparing humans and AI agents. *In Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3–6, 2011.* Proceedings 4 (pp. 122–132). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22887-2_13

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical image analysis, 5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage, 62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science, 367*(6473), eaaw4325. https://doi.org/10.1126/science.aaw4325

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, 103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., ... & Moser, M. B. (2008). Finite scale of spatial representation in the hippocampus. *Science, 321*(5885), 140–143. https://doi.org/10.1126/science.1157086

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron, 60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Maguire, E. A., Frackowiak, R. S., & Frith, C. D. (1997). Recalling routes around London: activation of the right hippocampus in taxi drivers. *The Journal of Neuroscience, 17*(18), 7103–7110. https://doi.org/10.1523/JNEUROSCI.17-18-07103.1997

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America, 97*(8), 4398–4403. https://doi.org/10.1073/PNAS.070039597

Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus, 16*(12), 1091–1101. https://doi.org/10.1002/HIPO.20233

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., ... & Kamitani, Y. (2008). Visual image reconstruction from human brain activity

using a combination of multiscale local image decoders. *Neuron, 60*(5), 915-929. https://doi.org/10.1016/j.neuron.2008.11.004

McCormick, C., St-Laurent, M., Ty, A., Valiante, T. A., & McAndrews, M. P. (2015). Functional and effective hippocampal–neocortical connectivity during construction and elaboration of autobiographical memory retrieval. *Cerebral cortex, 25*(5), 1297–1305. https://doi.org/10.1093/cercor/bht324

Mazzoni, G., Clark, A., De Bartolo, A., Guerrini, C., Nahouli, Z., Duzzi, D., ... & Venneri, A. (2019). Brain activation in highly superior autobiographical memory: The role of the precuneus in the autobiographical memory retrieval network. *Cortex, 120*, 588–602. https://doi.org/10.1016/j.cortex.2019.02.020

Namboodiri, V. M. K., & Stuber, G. D. (2021). The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron, 109*(22), 3552–3575. https://doi.org/10.1016/j.neuron.2021.09.034

Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., ... & Tank, D. W. (2021). Geometry of abstract learned knowledge in the hippocampus. *Nature, 595*(7865), 80–84. https://doi.org/10.1038/s41586-021-03652-7

O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research, 31*, 573–590. https://doi.org/10.1007/BF00239813

Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. Journal of *Neuroscience, 26*(17), 4596–4601. https://doi.org/10.1523/JNEUROSCI.1923-05.2006

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences, 17*(5), 230–240. https://doi.org/10.1016/j.tics.2013.03.005

Philippi, C. L., Tranel, D., Duff, M., & Rudrauf, D. (2015). Damage to the default mode network disrupts autobiographical memory retrieval. *Social Cognitive and Affective Neuroscience, 10*(3), 318–326. https://doi.org/10.1093/scan/nsu070

Peer, M., Ron, Y., Monsa, R., & Arzy, S. (2019). Processing of different spatial scales in the human brain. *Elife, 8*, e47492. https://doi.org/10.7554/eLife.47492

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map making: constructing, combining, and inferring on abstract cognitive maps. *Neuron, 107*(6), 1226–1238. https://doi.org/10.1016/j.neuron.2020.06.030

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Sciences, 25*(1), 37–54. https://doi.org/10.1016/j.tics.2020.10.004

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Robinson, J. L., Barron, D. S., Kirby, L. A., Bottenhorn, K. L., Hill, A. C., Murphy, J. E., ... & Fox, P. T. (2015). Neurofunctional topography of the human hippocampus. *Human Brain Mapping, 36*(12), 5018–5037. https://doi.org/10.1002/hbm.22936

Sullivan Giovanello, K., Schnyer, D. M., & Verfaellie, M. (2004). A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus, 14*(1), 5–8. https://doi.org/10.1002/hipo.10182

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia, 44*(12), 2189–2208. https://doi.org/10.1016/j.neuropsychologia.2006.05.023

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *Journal of Neuroscience, 31*(12), 4407–4420. https://doi.org/10.1523/JNEUROSCI.3335-10.2011

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron, 76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience, 15*(10), 655–669. https://doi.org/10.1038/nrn3785

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and space: Towards an understanding of the cognitive map. *Journal of Neuroscience, 35*(41), 13904–13911. https://doi.org/10.1523/JNEUROSCI.2614-15.2015

Schacter, D. L., & Madore, K. P. (2016). Remembering the past and imagining the future: Identifying and enhancing the contribution of episodic memory. *Memory Studies, 9*(3), 245–255. https://doi.org/10.1177/1750698016645231

Schwarb, H., Johnson, C. L., McGarry, M. D., & Cohen, N. J. (2016). Medial temporal lobe viscoelasticity and relational memory performance. *Neuroimage, 132*, 534–541. https://doi.org/10.1016/j.neuroimage.2016.02.067

Suárez, L. E., Markello, R. D., Betzel, R. F., & Misic, B. (2020). Linking structure and function in macroscale brain networks. *Trends in Cognitive Sciences, 24*(4), 302–315. https://doi.org/10.1016/j.tics.2020.01.004

Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., & Margulies, D. S. (2021). The default mode network in cognition: A topographical perspective. *Nature Reviews Neuroscience, 22*(8), 503–513. https://doi.org/10.1038/s41583-021-00480-4

Sorscher, B., Mel, G. C., Ocko, S. A., Giocomo, L. M., & Ganguli, S. (2023). A unified theory for the computational and mechanistic origins of grid cells. *Neuron, 111*(1), 121–137. https://doi.org/10.1016/j.neuron.2022.09.030

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience, 10*(2), 420–435. https://doi.org/10.1523/JNEUROSCI.10-02-00420.1990

Tartaglia, E. M., Clarke, A. M., & Herzog, M. H. (2017). What to choose next? a paradigm for testing human sequential decision making. *Frontiers in Psychology, 8*, 312. https://doi.org/10.3389/fpsyg.2017.00312

Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron, 96*(1), 228-241. https://doi.org/10.1016/j.neuron.2017.09.005

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage, 14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage, 21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., . . . Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage, 45*(1), S173–S186. https://doi.org/10.1016/j.neuroimage.2008.10.055

Wills, T. J., & Cacucci, F. (2014). The development of the hippocampal neural representation of space. *Current opinion in neurobiology, 24*, 111-119. https://doi.org/10.1016/j.conb.2013.09.006

Woolley, D. G., Mantini, D., Coxon, J. P., D'Hooge, R., Swinnen, S. P., Wenderoth, N. (2015). Virtual water maze learning in human increases functional connectivity between posterior hippocampus and dorsal caudate. *Human Brain Mapping, 36*(4), 1265–1277. https://doi.org/10.1002/hbm.22700

Wixted, J. T., Goldinger, S. D., Squire, L. R., Kuhn, J. R., Papesh, M. H., Smith, K. A., Steinmetz, P. N. (2018). Coding of episodic memory in the human

hippocampus. *Proceedings of the National Academy of Sciences, 115*(5), 1093–1098. https://doi.org/10.1073/pnas.1716443115

Weisberg, S. M., Newcombe, N. S., & Chatterjee, A. (2019). Everyday taxi drivers: Do better navigators have larger hippocampi?. *Cortex, 115*, 280–293. https://doi.org/10.1101/431155

Weisberg, S. M., & Ekstrom, A. D. (2021). Hippocampal volume and navigational ability: The map (ping) is not to scale. *Neuroscience & Biobehavioral Reviews, 126*, 102–112. https://doi.org/10.1016/j.neubiorev.2021.03.012

Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W., & Behrens, T. E. J. (2022). How to build a cognitive map. *Nature Neuroscience, 25*(10), 1257–1272. https://doi.org/10.1038/s41593-022-01153-y

Zeithamova, D., & Bowman, C. R. (2020). Generalization and the hippocampus: More than one story?. *Neurobiology of Learning and Memory, 175*, 107317. https://doi.org/10.1016/j.nlm.2020.107317

Zheng, X. Y., Hebart, M. N., Grill, F., Dolan, R. J., Doeller, C. F., Cools, R., & Garvert, M. M. (2024). Parallel cognitive maps for multiple knowledge structures in the hippocampal formation. *Cerebral Cortex*, bhad485. https://doi.org/10.1093/cercor/bhad485

Methods

Participants

Sixteen participants (mean age 23.00 ± 2.97, 9 female) with normal or corrected-to-normal vision were recruited from Tsinghua University. Informed consent was obtained from each participant before participation in accordance with the experimental protocol approved by the Institutional Review Board of Tsinghua University.

Relational Memory (RM) task

The experimental paradigm was a modified version of a sequential decision making task (Insa-Cabrera et al., 2011; Tartaglia et al., 2017). In each trial, a *Good* and a *Bad* cue were pseudo-ramdomly presented on two out of three horizontal positions of screen after a 10-s fixation. The position of the cues independently varies along 30 consecutive steps. The *Good* and *Bad* cue were never placed in the same position in sequence. Participants were required to predict the next position of the *Good* cue by placing a *Self* marker "avatar" at the anticipated location, and avoid the position of *Bad* cue during 1700 ms. After a self-paced response, feedback is immediately presented for 300 ms with the positive feedback indicated by a green upward arrow and negative feedback represented by a red downward arrow. If the chosen position is neither *Good* nor *Bad*, neutral feedback will be presented. Participants received +1, -1, and 0 virtual coin after positive, negative, and neutral feedback, respectively. If no response was made, the next step will be presented after 2000 ms. Participant's response, feedback, and reaction times were reported during experiment.

The experiment was conducted in two consecutive days. Participants performed the task in behavioral room on day 1 while they completed the task under MRI scanning on day 2, respectively.

The behavioral task employed a four-position design programmed by PsychoPy (version = 2022.1.0), where the length of sequences (i.e., numeric strings) varied from 2 to 8 steps with a step size of 1 and eight repetitions. This design resulted in 56 trials.

802 The total duration of experiment took approximately one hour. Experimental stimuli
803 were presented by a 15.4-inch MacBook Pro with a screen resolution of 2880 x 1800
804 pixels. Participants completed six practice trials prior to the experiment to ensure that
805 they fully understood the instructions.

806 The fMRI task was designed with three positions. The length of numeric strings
807 varied from 2 to 4 steps with a step size of 1 and eight repetitions, resulting in a total
808 of 24 trials, which were randomly assigned to 4 scanning sessions with 6 trials each
809 session. The total duration of fMRI experiment took approximately one hour. The
810 experimental stimuli were presented using a visual/audio stimulation system
811 (Shenzhen Sinorad SA-9939) with a customized-design (Shenzhen Sinorad Medical
812 Electronics Co., Ltd) MRI compatible 40" LED liquid crystal display. The distance
813 between participant and the display was 100 centimeters. The screen resolution was
814 1024 x 768 pixels. Participants viewed the stimuli with an angled mirror mounted on
815 the head coil.

816 **MRI Acquisition**

817 BOLD MRI images were acquired using a 3T Siemens Prisma scanner equipped with
818 a 64-channel phased-array head coil. Functional data were acquired with a
819 gradient-echo-planar T2* sequence (TR: 2000ms; TE: 34.0ms; matrix size:
820 100×100×72; flip angle: 90°; resolution: 2 × 2 × 2 mm^3; FoV = 200mm; number of
821 slices: 72; slice thickness: 2 mm; slice orientation: transversal). T1-weighted
822 three-dimensional anatomical images were collected to facilitate registration
823 (MPRAGE: TR: 2530 ms; TE: 2.27 ms; matrix size: 208×256×256; flip angle: 7°;
824 resolution: 1×1×1mm^3; number of slices: 208; slice thickness: 1 mm; slice orientation:
825 sagittal).

826 **fMRI Pre-processing**

827 T1-weighted structural and functional images were analyzed using FSL (FMRIB's
828 Software Library, version 6.0, http://fsl.fmrib.ox.ac.uk/fsl/; Woolrich et al., 2009;
829 Smith et al., 2004; Jenkinson et al., 2012). The non-brain tissues of high-resolution T1

830 images were removed using BET (Smith, 2002). The BOLD signal series of each
831 scanning session was preprocessed independently using FEAT (FSL's FMRI Expert
832 Analysis Tool, version 6.0; Woolrich et al., 2001; Woolrich et al., 2004), where
833 motion correction, slice-timing correction, and high-pass filtering at 100 s were
834 performed. For representational similarity analysis (RSA), BOLD signal series was
835 left unsmoothed to preserve fine-grained spatial information (Chadwick et al. 2012).
836 For functional connectivity (FC) analysis, BOLD signal series were smoothed using a
837 Gaussian kernel of full width at 5mm half-maximum (FWHM). These preprocessed
838 functional images of each participant were normalized to standard Montreal
839 Neurological Institute (MNI-152) template using FSL FLIRT (Jenkinson & Smith,
840 2001; Jenkinson et al., 2002) before being sent to group-level statistics.

841 **Representational similarity analysis**

842 The functional separation of the HPC was examined by the representational similarity
843 analysis (Kriegeskorte, Goebel, Bandettini, 2006; Kriegeskorte et al., 2008). First, the
844 BOLD signal series of task period were subdivided into two temporal intervals with
845 equal duration, which were labelled as the relational memory formation state and the
846 utilization state, respectively. Second, a general linear model (GLM) was created for
847 fitting the BOLD time series of each voxel, where 18 regressors were specified. These
848 regressors indicated the cognitive states corresponded by the rest, formation, and
849 utilization periods from each of the six trials, and the six motion parameters were
850 additionally included as covariates. Third, the multivoxel patterns were extracted from
851 the beta images using a 6mm radius sphere centered at each voxel of the HPC.
852 Pair-wise Pearson correlations were performed across the 18 cognitive states. This
853 procedure resulted in a 18 × 18 correlation matrix, which was Fisher's r-to-z
854 transformed. Fourth, two additional GLMs were created to examine the neural
855 representation of memory formation and utilization states, respectively. Specifically, a
856 binary regressor was used for each GLM to fit the correlation coefficients of cognitive
857 states, where 1 and 0 denoted “same” and “difference”, respectively. A higher beta
858 coefficient of the two GLMs in any voxel indicated stronger representational

859 similarity for the memory formation and utilization states. Then, the representational
860 similarity maps were transformed into MNI space, averaged across sessions, and
861 finally sent to second level analysis. Identical procedure was applied to the RSA of
862 the rest and task periods.

863 **Functional connectivity analysis**

864 To examine the connectivity separation of the HPC, we computed the connectivity
865 pattern of the HPC for each of the memory formation and memory utilization states.
866 Given the brain areas within the autobiographical network intrinsically connect to the
867 HPC (Conway et al., 2001; Spreng and Grady, 2010; Sestieri et al., 2011; Philippi et
868 al., 2015; Smallwood et al., 2021), we particularly focused on the connections
869 between the autobiographical network and the HPC. First, we derived the functional
870 mask of the autobiographical network by performing a contrast of the hippocampal
871 whole-brain connectivity map from the task period to the rest period (Fig. S2).
872 Specifically, a GLM was created to regress out the potential nuisance factors, which
873 included the six parameters derived from rigid body correction of head motion (Birn
874 et al., 2006; Wise et al., 2004), the mean time courses within the lateral ventricles, and
875 the mean time courses within the white matter (Bartels & Zeki, 2005; Fox et al., 2005).
876 Second, the BOLD signal series of the task and rest periods were extracted through
877 concatenating the TRs (i.e., 5 TRs) across trials (i.e., 6 trials) independently for each
878 session. Third, the whole brain connectivity map of the HPC for each of the rest and
879 task state was computed by performing Pearson correlations between the regional
880 time courses of the HPC (the HPC mask was obtained from the Harvard-Oxford (HO)
881 atlas included in FSL) and the BOLD signal series of each voxel of the remaining
882 brain. Fourth, the contrast in the connectivity of the HPC between the task period and
883 rest period was performed after Fisher's r-to-z transformation (Fig.S2). This result was
884 visualized using BrainNet Viewer (version 1.42, https://www.nitrc.org/projects/bnv).
885 The functional mask of the default mode network was then generated using the
886 contrast map with the voxels within the medial temporal lobe excluded. Finally, the
887 connectivity separation of the HPC was examined by computing the Pearson

correlations between the regional time course of default mode network and the time course of each voxel of the HPC in each of the memory formation and utilization states, which was extracted by concatenating the TRs (i.e., 15 TRs) across trials (i.e., 6 trials). The state-dependent connectivity maps of the HPC were transformed into standard MNI152 space, averaged across sessions after Fisher's r-to-z transformation, and sent to second level analysis.

Convolutional LSTM (C-LSTM) model

An encoder-decoder network programmed by PyTorch (version = 1.9; https://pytorch.org/) was built to perform the RM task, where the encoder included a convolutional layer followed by a fully connected layer, which projected the flattened output to a long short-term memory (LSTM) layer. The encoder accepted the images with dimension of 90×40×3 pixels as input and constructed the memory space storing the sequential relation of experimental stimuli. The decoder included a fully connected layer followed by a transposed convolution layer, which was used to generate visual output, indicating the predicted location of the *good* cue. In other words, the decoder utilized the sequential relations stored by memory space to guide actions on visual space. To examine the role of coarse / fine-grained scales akin to the hippocampal ventral-dorsal axis on relational memory formation and utilization, we modulated the visual resolution (VR), defined by 1/kernel size of convolutional layers (ranged from 1 to 15) and the memory resolution (MR), defined by the hidden size of LSTM (ranged from 1 to 15) to examine their influence on model performance. The kernel size of the convolutional layer was set the same as that of the deconvolutional layer based on the biological basis of the visual cortex (Miyawaki et al., 2008). In practice, the kernel size could vary independently, and the variance of the kernel size of the convolution layer had little effect on the findings (Fig. S3). During training, the Adam optimizer with a learning rate of 0.001 and a weight decay parameter of 1e-5 were used. The optimization process operates within a batch size of 8, which equals to the length of sequential images and spanned 1,000 epochs for each kernel size of convolutional layer and the hidden size of LSTM. The loss function is specified as

Mean Squared Error Loss. To improve the efficiency of network training, the encoder was independently trained in advance with a fully connected layer as decoder in predicting the upcoming location, then C-LSTM used the pretrained weights of encoder, which was fixed during training.

Statistics

The behavioral analysis was performed using Python 3.9 and SPSS 20. The second level analysis of RSA and functional connectivity analysis were performed using two tailed one-sample t-test. The generated t-maps used an initial threshold of $p = 0.01$. If no clusters were revealed, a liberal threshold of $p < 0.05$ was used. The reliability of the uncorrected clusters was further examined using a non-parametric statistical inference that does not make assumptions about the distribution of the data (Nichols & Holmes 2002). Specifically, 5,000 random sign-flips were performed on the parametric images of RSA and functional connectivity. The clusters with size higher than 95% of the maximal suprathreshold clusters from permutation distribution were reported. Same permutation procedure was applied to evaluate the statistical model fitness of participant behavior, the variation of the representational similarity and functional connectivity along hippocampal long axis, and the identification of unit type in C-LSTM. The main effect of C-LSTM parameter VR and MR and their interaction on task performance was examined by ANOVA. The unit type of C-LSTM was determined by computing the spatial frequency in both the horizontal and vertical dimension from the feature maps using Fast Fourier Transform (FFT). The linear relationship between the unit type and model performance was examined by Pearson correlation. ANOVA was employed to examine the significance of the FFT power variation of unit types relative to model performance.

Supplementary

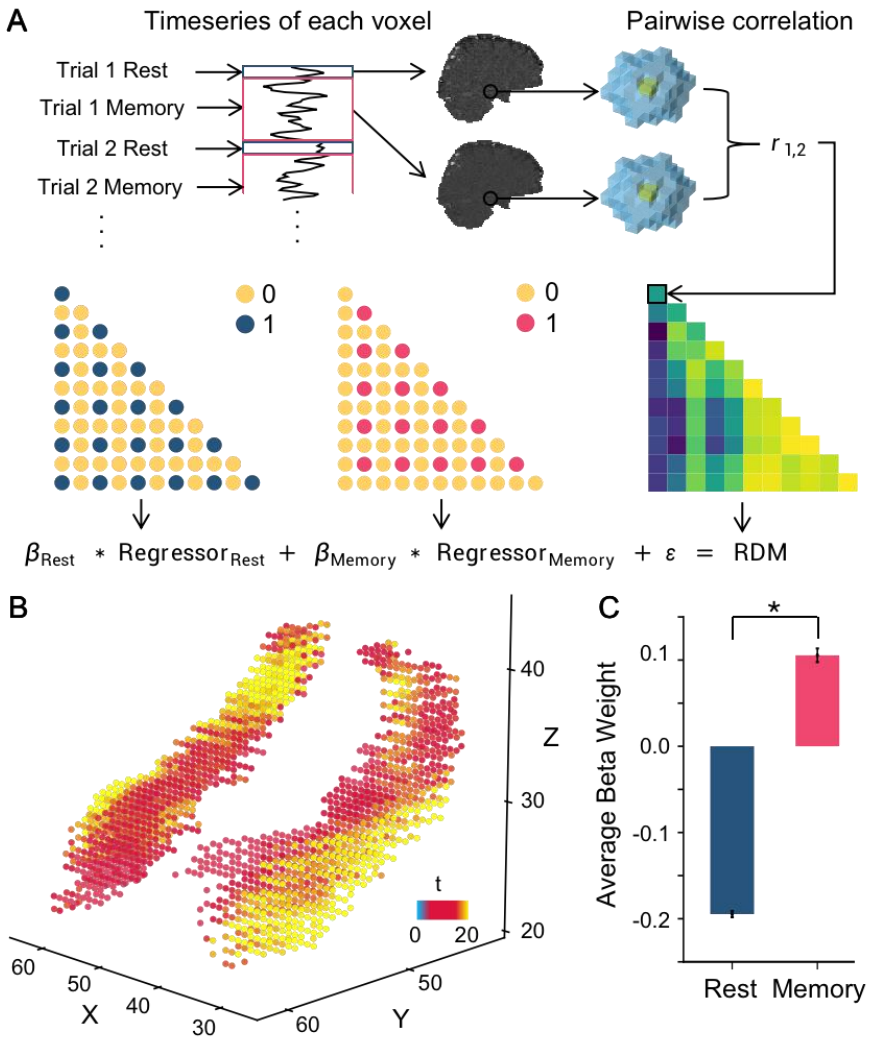

Figure S1. Representational similarity of rest and memory state. (A) The schema of RSA analysis. First, the multi-voxel patterns of the rest and memory state were derived by a GLM. Second, the 12x12 similarity matrix, built through the searchlight-based Pearson correlation of multi-voxel patterns among trials, was fitted by another GLM, which specified the cognitive state using binary regressors. The blue and red points in the binary matrix denote "same" between trial-pair. The estimated parameters of GLM were used for the analysis of (B). (B) The neural representation of memory relative to rest state. The value in each voxel represents t-value. The Y-axis corresponds to the coordinates of hippocampal long axis. (C) ROI analysis. The HPC showed significantly higher similarity in memory state relative to rest state ($t(15) = 33.45$; $p < 0.001$, two-tailed).* denotes the threshold of $p < 0.05$. Error bars denote standard error.

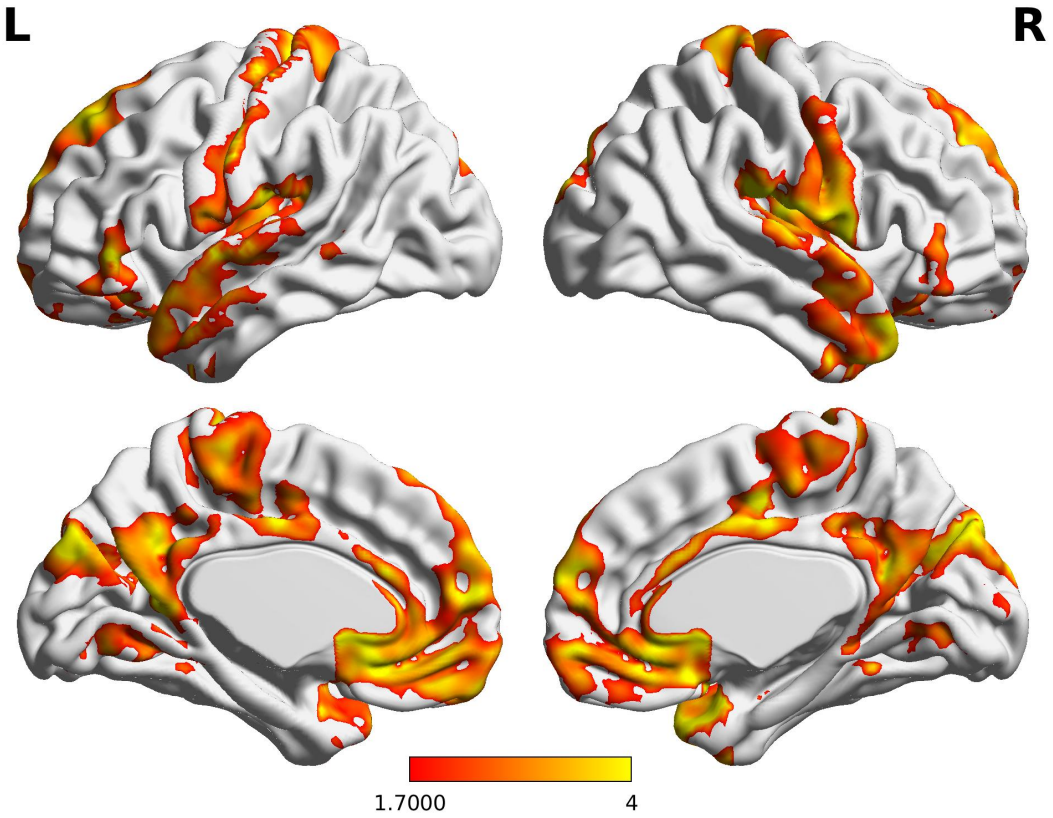

Figure S2. Functional connectivity of HPC in memory state compared to rest state. The functional connectivity was computed for each participant in both the memory and rest states. A Pearson correlation was performed between the mean BOLD signal series within the HPC and the voxel in the rest of brain. A contrast analysis was conducted comparing memory state relative to rest state. All clusters were corrected by False Discovery Rate (FDR) with a significance threshold of $p < 0.05$. colorbar denotes t value.

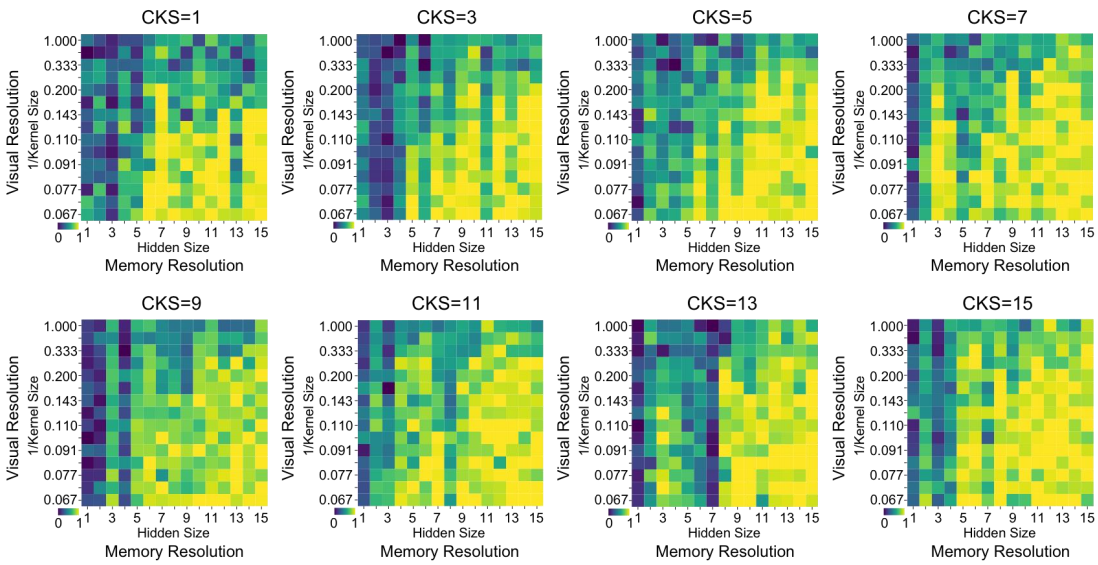

Figure S3. Behavioral performance of C-LSTM with fixed convolutional kernel size (CKS). Across a wide range of CKS values (i.e., varying from 1 to 15 in steps of 2), consistent patterns similar to Fig. 4B emerged. That is, optimal performance was achieved with smaller VR values (i.e., coarse representation during memory utilization) and larger MR values (i.e., fine-grained representation during memory formation). The color coding follows the scheme in Fig. 4B.